

The median of a random fuzzy number. The 1-norm distance approach

Abstract

In quantifying the central tendency of the distribution of a random fuzzy number (or fuzzy random variable in Puri and Ralescu’s sense), the most usual measure is the Aumann-type mean, which extends the mean of a real-valued random variable and preserves its main properties and behavior. Although such a behavior has very valuable and convenient implications, ‘extreme’ values or changes of data entail too much influence on the Aumann-type mean of a random fuzzy number. This strong influence motivates the search for a more robust central tendency measure. In this respect, this paper aims to explore the extension of the median to random fuzzy numbers. This extension is based on the 1-norm distance and its adequacy will be shown by analyzing its properties and comparing its robustness with that of the mean both theoretically and empirically.

Key words: 1-norm distance, fuzzy arithmetic, fuzzy numbers, median, random fuzzy numbers, statistical robustness

1 Introduction

In many real-life situations related to Social Sciences, Medicine, Engineering, etc., the information associated with some random experiments is imperfect. In this way, available data could correspond, for instance, to the valuation/rating of the employee productivity, costumer satisfaction, technological impact, agreement with a taken policy, quality of the soil, and so on.

As outlined and illustrated by Phillis and Kouikoglou [28], “fuzzy values are commonly used to express the way humans extract qualitative information from numerical, categorical or linguistic data, and the way they rate, summarize and process this information to make decisions and assessments.” Thus, the imprecision underlying many available data from surveys, ratings, etc. can be properly formalized in terms of fuzzy values and, in particular, fuzzy numbers.

The richness of the scale of fuzzy numbers (including real and interval values as special elements) allows us to cope with a wide set of imprecise data, as those mentioned above. Instead of modeling the type of data by means of either numerical or categorical data, which would be less accurate or expressive, the

fuzzy scale integrates the manageability and diversity/variability of the numerical scale and the interpretability and ability to capture the imprecision of the categorical scale. Furthermore, fuzzy numbers become a flexible and easy-to-use tool which enables us to exploit the subjectivity that is often involved in perceiving and expressing the available information. They have a very intuitive meaning and potential users can friendly understand the required basic notions and ideas to manage fuzzy data.

The concept of random fuzzy number, or more generally that of random fuzzy set (formerly called fuzzy random variables) in Puri and Ralescu's sense [31], has been introduced as a mathematical model for data generation processes associating fuzzy values with the outcomes of random experiments. Probabilistic aspects of random fuzzy sets have received a lot of attention in the literature since more than two decades. Most of the studies have concerned the measurability or limit theorems for sequences of these random elements (see, for instance, Colubi *et al.* [5], [3], [4], Molchanov [25], Proske and Puri [29], Li and Ogura [23]).

However, the statistical aspects have not been examined to the same extent. In developing statistics with random fuzzy sets some distinctive features in contrast to the statistics with random variables become crucial, namely,

- the usual arithmetic with fuzzy values determines a semilinear structure, which implies that there is no generally applicable definition for the difference of fuzzy values preserving the connection with the sum in the numerical case;
- the lack of a universally acceptable total ordering between fuzzy values (although several ones have been proposed, like the well-known by Yager [37]; in many situations these proposals lead to rather plausible rankings, but the conclusions are not that clear or acceptable in some other cases);
- the lack of realistic and easy-to-handle 'parametric' families for the distribution of random fuzzy sets.

Most of the statistical developments with random fuzzy sets, which have been carried out in the last decade (see, for instance, Körner [21], Montenegro *et al.* [26], [27], Gil *et al.* [14], and González-Rodríguez *et al.* [16], [17]) refer to the so-called Aumann-type mean of a random fuzzy set in Puri and Ralescu's sense [31]. The lack of a well-defined difference between fuzzy values preserving the connection with the sum in the numerical case has been overcome by using an L^2 type distance between fuzzy values with similar interpretation and mission as the Euclidean distance for numerical data. Actually, this is one of the key advantages of dealing with fuzzy data instead of categorical data.

The Aumann-type mean of a random fuzzy number in accordance with Puri and Ralescu shares the main skills and worth of the means of random elements. Additional implications for the statistical analysis of real-valued data have been carried out in González-Rodríguez *et al.* [15]. Consequently, the crucial

role played by, as well as the convenient results around, the Aumann-type of a random fuzzy number is not at all a matter for discussion. Nevertheless, in summarizing or representing the centrality of a random fuzzy number it should be emphasized that, also as for the real-valued case, the mean value is very sensitive to the change of values or to the existence of ‘extreme’ values; moreover, the intrinsic imprecision can add an extra sensitivity.

For purposes of introducing a more robust summary measure of the central tendency for random fuzzy numbers, reducing the influence of ‘extreme’ values (outliers) or the changes in values, one can explore the extension of the median of a random variable. The lack of a universally acceptable total ordering between fuzzy numbers will be overcome in the approach followed in this paper by using an L^1 type distance between fuzzy numbers based on the 1-norm and on the infimum/supremum representation of the levels (or, equivalently, on the support function) of the fuzzy numbers. It should be made clear that by no means we aim to introduce a measure to compete with the Aumann-type mean value, but just to define a measure offering a very convenient representation and description of the central tendency in case of asymmetric sample or population fuzzy data, and hence enriching in some respects the statistical analysis of these data.

In Section 2 the preliminaries on fuzzy numbers, arithmetic and metric between them, and the concepts of random fuzzy number and Aumann-type mean will be recalled or established. Section 3 presents a simulation study illustrating the sensitivity of the mean value and, hence, motivating the need for a more robust summary measure for central tendency. Section 4 contains the suggested approach to extend the median, and states a convention for the practical computation of the median, which will be illustrated by means of a real-life example. Section 5 shows how the main properties of the median in the real-valued case are preserved when dealing with fuzzy data. Section 6 discusses some relevant properties (consistency and robustness) of the sample median as an estimator of the population median. The robustness of the sample mean and sample median as estimators of their population counterparts will be also compared both theoretically and empirically. Finally, some future research directions will be commented.

2 Preliminaries

A (bounded) *fuzzy number* (in the literature sometimes also referred to as a compact fuzzy interval) is formalized as a mapping $\tilde{U} : \mathbb{R} \rightarrow [0, 1]$ so that for each $\alpha \in [0, 1]$ the α -level set, $\tilde{U}_\alpha = \{x \in \mathbb{R} : \tilde{U}(x) \geq \alpha\}$ for $\alpha > 0$ and $\tilde{U}_0 = \text{cl}\{x \in \mathbb{R} : \tilde{U}(x) > 0\}$, is a nonempty, closed and bounded interval. For each $x \in \mathbb{R}$, the value $\tilde{U}(x)$ is interpreted as the ‘degree of compatibility of x with the property characterizing \tilde{U} ’ or the ‘degree of possibility of the assertion “ x is \tilde{U} ”’. The space of (bounded) fuzzy numbers which will model

data in the paper will be denoted by $\mathcal{F}_c^*(\mathbb{R})$.

The statistical analysis of fuzzy data requires as operators the sum and the product by a scalar. These operators are assumed to be based on the usual fuzzy arithmetic applying Zadeh's extension principle [38]. The two operations can be equivalently formalized as the level-wise extensions of the usual interval-valued operations. That is, given two fuzzy numbers $\tilde{U}, \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})$ and a real number γ , the *sum of \tilde{U} and \tilde{V}* is defined as $\tilde{U} + \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})$ such that for each $\alpha \in [0, 1]$

$$(\tilde{U} + \tilde{V})_\alpha = \text{Minkowski sum of } \tilde{U}_\alpha \text{ and } \tilde{V}_\alpha = \{y + z : y \in \tilde{U}_\alpha, z \in \tilde{V}_\alpha\},$$

and the *product of \tilde{U} by the scalar γ* is defined as $\gamma \cdot \tilde{U} \in \mathcal{F}_c^*(\mathbb{R})$ such that for each $\alpha \in [0, 1]$

$$(\gamma \cdot \tilde{U})_\alpha = \gamma \cdot \tilde{U}_\alpha = \{\gamma \cdot y : y \in \tilde{U}_\alpha\}.$$

As pointed out before, when the space $\mathcal{F}_c^*(\mathbb{R})$ is endowed with these two operations it does not have a linear but a semilinear (actually a conical) structure. This fact requires that we have to be very careful in handling fuzzy data, because it is not possible to establish a well-defined difference between fuzzy numbers that keeps the connection with the sum in the numerical case (see Bouchon-Meunier *et al.* [1]). More precisely, although a well-defined difference could be simply established as $\tilde{U} - \tilde{V} = \tilde{U} + (-1) \cdot \tilde{V}$, it turns out that in general $\tilde{U} - \tilde{V} + \tilde{V} \neq \tilde{U}$. On the other hand, if the difference is established to ensure that the last equality holds (i.e., if the Hukuhara difference [19] is level-wise considered), then the operation is not well-defined for most of the fuzzy numbers.

The arithmetic with fuzzy numbers does not coincide directly with the usual arithmetic with mappings although, as outlined in González-Rodríguez *et al.* [16], a correspondence can be established whenever fuzzy numbers are identified with their support function. Puri and Ralescu [30] have defined the support function of a fuzzy value by extending the notion of the support function of a set by Castaing and Valadier [2] level-wise. In the particular case of $\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})$ the *support function* of \tilde{U} is the mapping $s_{\tilde{U}} : \{-1, 1\} \times (0, 1] \rightarrow \mathbb{R}$ defined so that

$$\tilde{U}_\alpha = [\inf \tilde{U}_\alpha, \sup \tilde{U}_\alpha] = [-s_{\tilde{U}}(-1, \alpha), s_{\tilde{U}}(1, \alpha)],$$

that is, it corresponds to the inf / sup characterization of the levels of the fuzzy numbers.

Because of the lack of a general suitable definition for the difference of two fuzzy numbers, in the statistical developments with random fuzzy numbers a crucial alternate role is played by the metrics between fuzzy values. Among the best known metrics between fuzzy values one can remark because of several valuable properties the ones stated by Klement *et al.* [20] and being based on Hausdorff distance between convex compact sets, as well as the ones stated by

Diamond and Kloeden [8]. Also some L^2 -metrics have been shown to be relevant because of the Fréchet principle (see, for instance, Körner and Näter [22] or recently Trutschnig *et al.* [34]) in connection with statistical developments regarding the fuzzy mean value.

Similar arguments suggest that L^1 -metrics would be relevant when dealing with an extension of the median. In particular, an easy-to-use L^1 -distance between fuzzy numbers that allows to establish a Rådström type isometrical embedding (see Diamond and Kloeden [8]) will now be recalled:

Definition 2.1 *The mapping $\rho_1 : \mathcal{F}_c^*(\mathbb{R}) \times \mathcal{F}_c^*(\mathbb{R}) \rightarrow [0, +\infty)$ such that for $\tilde{U}, \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})$*

$$\rho_1(\tilde{U}, \tilde{V}) = \|s_{\tilde{U}} - s_{\tilde{V}}\|_1 = \frac{1}{2} \int_{(0,1]} \left(|\inf \tilde{U}_\alpha - \inf \tilde{V}_\alpha| + |\sup \tilde{U}_\alpha - \sup \tilde{V}_\alpha| \right) d\alpha$$

*will be called the **1-norm distance between fuzzy numbers**.*

The distance ρ_1 will be shown in Section 4 to be easy-to-handle for purposes of extending the notion of the median. Furthermore, other interesting results are satisfied in connection with ρ_1 (see Diamond and Kloeden [8]), namely,

Proposition 2.1 *The metric ρ_1 on $\mathcal{F}_c^*(\mathbb{R}) \times \mathcal{F}_c^*(\mathbb{R})$ is topologically equivalent to the metric d_1 by Klement et al. [20] on $\mathcal{F}_c^*(\mathbb{R}) \times \mathcal{F}_c^*(\mathbb{R})$, which is given by*

$$d_1(\tilde{U}, \tilde{V}) = \int_{(0,1]} d_H(\tilde{U}_\alpha, \tilde{V}_\alpha) d\alpha,$$

with d_H being the well-known Hausdorff metric on the space of nonempty closed and bounded intervals. More precisely,

$$\frac{d_1(\tilde{U}, \tilde{V})}{2} \leq \rho_1(\tilde{U}, \tilde{V}) \leq d_1(\tilde{U}, \tilde{V})$$

for all $\tilde{U}, \tilde{V} \in \mathcal{F}_c^(\mathbb{R})$. As a consequence, $(\mathcal{F}_c^*(\mathbb{R}), \rho_1)$ is a separable metric space.*

A Rådström type isometrical embedding of $\mathcal{F}_c^*(\mathbb{R})$ onto a convex cone of a Hilbert space of functions with the functional arithmetic and the ρ_1 metric can be established by following the results in Puri and Ralescu [30] and Klement *et al.* [20].

Proposition 2.2 *The mapping $s : \mathcal{F}_c^*(\mathbb{R}) \rightarrow \mathcal{H}_1^*$ = space of the L^1 -type real-valued functions defined on $\{-1, 1\} \times (0, 1]$ such that $s(\tilde{U}) = s_{\tilde{U}}$ for all $\tilde{U} \in \mathcal{F}_c^*(\mathbb{R}) = \{\tilde{U} \in \mathcal{F}_c^*(\mathbb{R}) : s_{\tilde{U}} \in \mathcal{H}_1^*\}$ preserves the semilinear structure of $\mathcal{F}_c^*(\mathbb{R})$, and states an isometrical embedding of $\mathcal{F}_c^*(\mathbb{R})$ with the fuzzy arithmetic and*

the metric ρ_1 onto a closed convex cone of \mathcal{H}_1^* with the functional arithmetic and the metric based on the 1-norm.

As an immediate implication from the preceding isometry, a fuzzy number $\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})$ can be ‘identified’ with the functional value $s_{\tilde{U}} \in \mathcal{H}_1^*$, the fuzzy arithmetic can be ‘identified’ with the functional one, since $s_{\tilde{U}+\tilde{V}} = s_{\tilde{U}} + s_{\tilde{V}}$ and $s_{\gamma \cdot \tilde{U}} = \gamma \cdot s_{\tilde{U}}$, and (as defined) $\rho_1(\tilde{U}, \tilde{V}) = \|s_{\tilde{U}} - s_{\tilde{V}}\|_1$.

Random fuzzy numbers have been introduced as random elements taking on fuzzy numbered values. They model random mechanisms that produce fuzzy data. The definition by Puri and Ralescu [31] can be equivalently formalized in different ways (see, for instance, Colubi *et al.* [3], González-Rodríguez *et al.* [16]). Given a probability space (Ω, \mathcal{A}, P) that models the considered random experiment, an associated *random fuzzy number* (for short *RFN*) is a mapping $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c^*(\mathbb{R})$ such that for all $\alpha \in (0, 1]$ the α -level mapping \mathcal{X}_α is a compact random interval (that is, for all $\alpha \in (0, 1]$ the real-valued mappings $\inf \mathcal{X}_\alpha$ and $\sup \mathcal{X}_\alpha$ are random variables.)

Based on Colubi *et al.* [3], if $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c^*(\mathbb{R})$ is an RFN, then it is a Borel-measurable mapping with respect to the Borel σ -field generated on $\mathcal{F}_c^*(\mathbb{R})$ by the topology associated with ρ_1 . The Borel measurability will enable to consider trivially the *induced distribution of an RFN* as well as the *independence of RFNs*.

The *Aumann-type mean of an RFN* has been defined by Puri and Ralescu [31] as the fuzzy number $\tilde{E}(\mathcal{X}) \in \mathcal{F}_c^*(\mathbb{R})$ such that for all $\alpha \in (0, 1]$

$$\left(\tilde{E}(\mathcal{X})\right)_\alpha = [E(\inf \mathcal{X}_\alpha), E(\sup \mathcal{X}_\alpha)].$$

Equivalently, it can be formalized as the fuzzy number $\tilde{E}(\mathcal{X}) \in \mathcal{F}_c^*(\mathbb{R})$ such that $s_{\tilde{E}(\mathcal{X})} = E(s_{\mathcal{X}})$ whenever the involved real-valued (or functional-valued) expectations exist.

\tilde{E} preserves all the main properties of the mean of a random variable. Thus, it is equivariant under ‘linear’ transformations and under the sum of RFS’s, it is coherent with the usual fuzzy arithmetic, and it is the ‘Fréchet expectation’ of \mathcal{X} w.r.t. several L^2 -type metrics. Moreover, it is also supported by several Strong Laws of Large Numbers w.r.t. most of the metrics we can consider.

3 Motivating the extension of the median for random fuzzy numbers

The Aumann-type mean value is the most common candidate to get some idea about the central tendency of a sample or population of fuzzy data. Neverthe-

less, one should know that in addition to preserving the properties indicated at the end of the last section, the Aumann-type fuzzy mean also inherits from the real-valued case the sensitivity of the mean to either perturbations in data or the existence of extreme values (frequently referred to as outliers).

To illustrate in which way contamination affects the Aumann-type fuzzy mean some simulations are now considered. In these simulation studies trapezoidal-valued RFNs will be considered, each of them being characterized by the following four real-valued random variables:

- $X_1 = (\inf \mathcal{X}_1 + \sup \mathcal{X}_1)/2$, $X_2 = (\sup \mathcal{X}_1 - \inf \mathcal{X}_1)/2$,
- $X_3 = \inf(\mathcal{X}_1) - \inf(\mathcal{X}_0)$, $X_4 = \sup(\mathcal{X}_0) - \sup(\mathcal{X}_1)$,

that is, $\mathcal{X} = \text{Tra}(X_1 - X_2 - X_3, X_1 - X_2, X_1 + X_2, X_1 + X_2 + X_4)$, the last three ones being nonnegative random variables.

In the simulations, the population mean will be approximated by a Monte Carlo approach from a sample of size $n = 100000$ which is assumed to be split into a subsample of size $n(1 - c_p)$ associated with a noncontaminated distribution and a subsample of size $n c_p$ associated with a contaminated one, i.e., c_p denotes the proportion of contamination. An additional element to control contamination will be given by C_D , which measures (in terms of percentages) how far the distribution of the contaminated subsample is from the distribution of the noncontaminated one. To determine the effect of the contamination on the mean of the RFN \mathcal{X} , the expected distance between the noncontaminated distribution and the approximated mean is collected in Table 1; for this purpose, we have considered the ρ_1 metric as well as the ρ_2 metric, which is defined (cf. Diamond and Kloeden [8]) as

$$\rho_2(\tilde{U}, \tilde{V}) = \sqrt{\frac{1}{2} \int_{(0,1]} \left([\inf \tilde{U}_\alpha - \inf \tilde{V}_\alpha]^2 + [\sup \tilde{U}_\alpha - \sup \tilde{V}_\alpha]^2 \right) d\alpha}$$

since the Aumann-type mean is the Fréchet expectation of the RFN w.r.t. ρ_2 .

Some situations have been simulated for different values of c_p and C_D in two cases, namely, one in which random variables X_i are independent (Case 1) and another one in which they are dependent (Case 2). More specifically, Case 1 will assume that

- $X_1 \rightsquigarrow \mathcal{N}(0, 1)$ and $X_2, X_3, X_4 \equiv \chi_1^2$ for the non contaminated subsample,
- $X_1 \rightsquigarrow \mathcal{N}(0, 3) + C_D$ and $X_2, X_3, X_4 \equiv \chi_4^2 + C_D$ for the contaminated subsample,

whereas Case 2 will assume that

- $X_1 \rightsquigarrow \mathcal{N}(0, 1)$ and $X_2, X_3, X_4 \equiv 1/(X_1^2 + 1)^2 + .1 \cdot \chi_1^2$ for the non contaminated subsample,
- $X_1 \rightsquigarrow \mathcal{N}(0, 3) + C_D$ and $X_2, X_3, X_4 \equiv 1/(X_1^2 + 1)^2 + .1 \cdot \chi_1^2 + C_D$ for the contaminated subsample.

c_P	c_D	Case 1 (ρ_1)	Case 1 (ρ_2)	Case 2 (ρ_1)	Case 2 (ρ_2)
0.0	0	1.411700	1.591844	1.409091	1.621120
0.0	1	1.405566	1.585511	1.409232	1.621485
0.0	5	1.412428	1.592913	1.410245	1.624613
0.0	10	1.404929	1.584392	1.410163	1.623303
0.0	100	1.412076	1.592221	1.404439	1.616426
0.1	0	1.524461	1.724165	1.402352	1.613895
0.1	1	1.581919	1.787890	1.419247	1.633259
0.1	5	1.918657	2.158065	1.657168	1.899890
0.1	10	2.482627	2.798702	2.132237	2.449560
0.1	100	15.533895	18.696102	15.032214	18.254039
0.2	0	1.717045	1.930706	1.401529	1.612721
0.2	1	1.893293	2.116487	1.458557	1.676867
0.2	5	2.813724	3.135933	2.107559	2.424991
0.2	10	4.165025	4.722812	3.327517	3.896056
0.2	100	30.927565	37.309440	29.903652	36.424357
0.4	0	2.273397	2.497495	1.380048	1.586503
0.4	1	2.751243	2.997065	1.539542	1.769052
0.4	5	4.960736	5.502123	3.242665	3.809713
0.4	10	7.897286	9.010965	5.998302	7.214740
0.4	100	61.812702	74.597288	59.725347	72.798039

Table 1. Mean distances of the mixed (partially contaminated and noncontaminated) sample Aumann-type mean to the noncontaminated distribution of an RFN

On the basis of these simulations one can empirically conclude from Table 1 that the higher the perturbation, the worse the sample mean summarizes the noncontaminated distribution, with the influence of the contamination being quite strong. This strong influence will be substantially reduced by considering a new centrality measure allowing us to achieve a higher robustness, as will be shown in the next sections.

4 The ρ_1 median of a random fuzzy number

The strong influence of changes or the existence of ‘extreme’ values illustrated in Section 3 motivates the introduction of a more robust central tendency measure extending the notion of median to random fuzzy numbers. The median of a real-valued random variable is usually defined in two equivalent ways, namely: either as a ‘middle position’ value with respect to a specified ranking, or as a value minimizing the mean distance to the distribution of the variable through an L^1 -type metric. Since fuzzy numbers cannot be ranked through

a universally acceptable total ordering, we will consider the extension of the second way based on the metric ρ_1 . Thus,

Definition 4.1 *Given a probability space (Ω, \mathcal{A}, P) and an associated RFN \mathcal{X} , the **median** (or the medians) of the distribution of \mathcal{X} is the fuzzy number (or fuzzy numbers) $\widetilde{\text{Me}}(\mathcal{X}) \in \mathcal{F}_c^*(\mathbb{R})$ such that*

$$E\left(\rho_1(\mathcal{X}, \widetilde{\text{Me}}(\mathcal{X}))\right) = \min_{\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})} E\left(\rho_1(\mathcal{X}, \tilde{U})\right),$$

whenever these expectations exist.

Consequently, $\widetilde{\text{Me}}(\mathcal{X})$ is any fuzzy number that minimizes the mean ρ_1 -distance between a fuzzy number and the distribution of the RFN, which corroborates the fact that the median is a central tendency measure.

Two key questions at this stage are whether such fuzzy number-valued median exists and whether it can be computed easily in practice. The next result guarantees that at least one such median always exists.

Theorem 4.1 *Given a probability space (Ω, \mathcal{A}, P) and an associated RFN \mathcal{X} , for any $\alpha \in (0, 1]$ we have that the fuzzy number $\widetilde{\text{Me}}(\mathcal{X}) \in \mathcal{F}_c^*(\mathbb{R})$ such that*

$$\left(\widetilde{\text{Me}}(\mathcal{X})\right)_\alpha = [\text{Me}(\inf \mathcal{X}_\alpha), \text{Me}(\sup \mathcal{X}_\alpha)],$$

where in case $\text{Me}(\inf \mathcal{X}_\alpha)$ or $\text{Me}(\sup \mathcal{X}_\alpha)$ are nonunique the most usual convention will be followed:

- $\text{Me}(\inf \mathcal{X}_\alpha)$ will be chosen to be the midpoint of the interval of medians of $\inf \mathcal{X}_\alpha$,
- $\text{Me}(\sup \mathcal{X}_\alpha)$ will be chosen to be the midpoint of the interval of medians of $\sup \mathcal{X}_\alpha$,

is a median of the distribution of \mathcal{X} in accordance with Definition 4.1.

Proof. Indeed,

- ▷ On one hand, whatever $\alpha \in (0, 1]$ and $\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})$ may be, since $\inf \tilde{U}_\alpha, \sup \tilde{U}_\alpha \in \mathbb{R}$, and $\inf \mathcal{X}_\alpha$ and $\sup \mathcal{X}_\alpha$ are random variables, we have that

$$\begin{aligned} E[|\inf \mathcal{X}_\alpha - \text{Me}(\inf \mathcal{X}_\alpha)|] &\leq E[|\inf \mathcal{X}_\alpha - \inf \tilde{U}_\alpha|], \\ E[|\sup \mathcal{X}_\alpha - \text{Me}(\sup \mathcal{X}_\alpha)|] &\leq E[|\sup \mathcal{X}_\alpha - \sup \tilde{U}_\alpha|], \end{aligned}$$

whence

$$\begin{aligned} &E\left(\rho_1(\mathcal{X}, \tilde{U})\right) \\ &= \frac{1}{2} \int_{(0,1]} E[|\inf \mathcal{X}_\alpha - \inf \tilde{U}_\alpha|] d\alpha + \frac{1}{2} \int_{(0,1]} E[|\sup \mathcal{X}_\alpha - \sup \tilde{U}_\alpha|] d\alpha \\ &\geq \frac{1}{2} \int_{(0,1]} E[|\inf \mathcal{X}_\alpha - \text{Me}(\inf \mathcal{X}_\alpha)|] d\alpha + \frac{1}{2} \int_{(0,1]} E[|\sup \mathcal{X}_\alpha - \text{Me}(\sup \mathcal{X}_\alpha)|] d\alpha \\ &= E\left(\rho_1(\mathcal{X}, \widetilde{\text{Me}}(\mathcal{X}))\right). \end{aligned}$$

▷ On the other hand, intervals $[\text{Me}(\inf \mathcal{X}_\alpha), \text{Me}(\sup \mathcal{X}_\alpha)]$ correspond to the α -levels of a fuzzy number. Thus, for any $\alpha \in (0, 1]$ they are well-defined intervals, because of the considered convention $\inf \mathcal{X}_\alpha \leq \sup \mathcal{X}_\alpha$ entails that $\text{Me}(\inf \mathcal{X}_\alpha) \leq \text{Me}(\sup \mathcal{X}_\alpha)$. Moreover, $\text{Me}(\inf \mathcal{X}_1) \leq \text{Me}(\sup \mathcal{X}_1)$ ensures that the 1-level is nonempty.

Since $\inf \mathcal{X}_\alpha$ ($\sup \mathcal{X}_\alpha$) is a nondecreasing (respectively, a nonincreasing) function of α , then $\text{Me}(\inf \mathcal{X}_\alpha)$ is also nondecreasing (respectively, nonincreasing).

To conclude one should verify that $\text{Me}(\inf \mathcal{X}_\alpha)$ and $\text{Me}(\sup \mathcal{X}_\alpha)$ are left-continuous at every $\alpha \in (0, 1]$. If $\{\alpha_n\}_n \uparrow \alpha \in (0, 1]$ as $n \rightarrow \infty$, then for all element in Ω we have that $\{\inf \mathcal{X}_{\alpha_n}\}_n \uparrow \inf \mathcal{X}_\alpha$ and because of the considered convention the sequence $\{\text{Me}(\inf \mathcal{X}_{\alpha_n})\}_n \uparrow$ is bounded above, $\text{Me}(\inf \mathcal{X}_\alpha)$ being an upper bound. Hence, a limit for this sequence exists and will be denoted by $L_\alpha = \lim_{n \rightarrow \infty} \text{Me}(\inf \mathcal{X}_{\alpha_n}) \leq \text{Me}(\inf \mathcal{X}_\alpha)$.

$L_\alpha = \text{Me}(\inf \mathcal{X}_\alpha)$, since for all $\omega \in \Omega$ we have that

$$0.5 \leq P(\inf \mathcal{X}_{\alpha_n} \leq \text{Me}(\inf \mathcal{X}_{\alpha_n})) \leq P(\inf \mathcal{X}_{\alpha_n} \leq L_\alpha)$$

and

$$\{(\inf \mathcal{X}_{\alpha_n} \leq L_\alpha)\}_n \downarrow \bigcap_n (\inf \mathcal{X}_{\alpha_n} \leq L_\alpha) = (\inf \mathcal{X}_\alpha \leq L_\alpha),$$

whence

$$P(\inf \mathcal{X}_\alpha \leq L_\alpha) = P\left(\lim_n (\inf \mathcal{X}_{\alpha_n} \leq L_\alpha)\right) = \lim_{n \rightarrow \infty} P(\inf \mathcal{X}_{\alpha_n} \leq L_\alpha) \geq 0.5.$$

Following similar arguments,

$$\begin{aligned} P(\inf \mathcal{X}_\alpha < L_\alpha) &= P\left(\bigcup_n (\inf \mathcal{X}_\alpha < \text{Me}(\inf \mathcal{X}_{\alpha_n}))\right) \\ &= P\left(\lim_n (\inf \mathcal{X}_\alpha < \text{Me}(\inf \mathcal{X}_{\alpha_n}))\right) = \lim_{n \rightarrow \infty} P(\inf \mathcal{X}_\alpha < \text{Me}(\inf \mathcal{X}_{\alpha_n})) \\ &\leq \lim_{n \rightarrow \infty} P(\inf \mathcal{X}_{\alpha_n} < \text{Me}(\inf \mathcal{X}_{\alpha_n})) \leq 0.5. \end{aligned}$$

Consequently, taking into account the considered convention, we have that $L_\alpha \geq \text{Me}(\inf \mathcal{X}_\alpha)$ and, therefore, $L_\alpha = \text{Me}(\inf \mathcal{X}_\alpha)$.

Analogously, if $\{\alpha_n\}_n \uparrow \alpha \in (0, 1]$ as $n \rightarrow \infty$, it holds that $\{\sup \mathcal{X}_{\alpha_n}\}_n \downarrow \sup \mathcal{X}_\alpha$ and the sequence $\{\text{Me}(\sup \mathcal{X}_{\alpha_n})\}_n \downarrow$ and it is bounded below by $\text{Me}(\sup \mathcal{X}_\alpha)$ so that there exists $L'_\alpha = \lim_{n \rightarrow \infty} \text{Me}(\sup \mathcal{X}_{\alpha_n})$ and we can easily prove that $L'_\alpha = \text{Me}(\sup \mathcal{X}_\alpha)$. \square

Remark 4.1 It should be pointed out that with the convention in Theorem 4.1 it is easy to compute a fuzzy numbered solution of Definition 4.1. However, if we do not consider some valid conventions, then the result can fail. That is, in case $\text{Me}(\inf \mathcal{X}_\alpha)$ or $\text{Me}(\sup \mathcal{X}_\alpha)$ are nonunique, there are choices for them which do not determine a fuzzy number. As a counterexample, consider the RFN \mathcal{X} taking on the triangular values $\tilde{x}_1 = \text{Tra}(0, 1, 1, 2)$ and $\tilde{x}_2 = \text{Tra}(1, 2, 2, 3)$ both with induced probability .5; then, for $\alpha = .75$ we have that $\text{Me}(\inf \mathcal{X}_{.75})$ is any value in $[.75, 1.75]$, whereas $\text{Me}(\sup \mathcal{X}_{.75})$ is any value in

[1.25, 2.25], so that some choices for the medians of $\inf \mathcal{X}_{.75}$ and $\sup \mathcal{X}_{.75}$ would lead to empty α -levels. To avoid an unnecessary cumbersome checking and to ease the study of the properties of the median, from now on the median will be assumed to be defined as the unique fuzzy number in Theorem 4.1.

Remark 4.2 In contrast to the median for random variables, the median of an RFN as introduced in this paper does not necessarily correspond to one of the values of the RFN. As an example corroborating this assertion and illustrating the computation of the median we consider the RFN associated with the ‘overall rating’ of a course on the sample/population of 27 students for which values have been gathered as shown in Table 2. The corresponding median has been approximated by using a large number of levels, following ideas similar to those by Trutschnig and Lubiano [35], and is graphically displayed in Figure 3.

A real-life example is now considered to illustrate the computation of the median of an RFN as well as the comments in the last remark.

Example 4.1 In most of academic institutions it is a common practice to perform surveys among students to evaluate their satisfaction or to rate the level of different courses which are delivered at them. For this purpose questionnaires are designed to gather their students’ opinions and judgements. Most of these questionnaires are based on a pre-specified response format, often related to a Likert scale (like, for instance, the one including as possible responses VERY HIGH level, RATHER HIGH level, HIGH level, SOMEWHAT HIGH level, and so on). For the statistical analysis of the responses, these are treated either as categorical (for which statistical methods are rather limited) or coded by and handled as integer numbers (integer coding usually not reflecting the real differences between distinct values, and not capturing the imprecision and subjectivity which is intrinsic to these responses).

In several studies (see, for instance, González-Rodríguez *et al.* [16]), it has been suggested to use instead of Likert or integer scales, whenever it is reasonable and feasible, the scale of fuzzy numbers. This scale enables us to reflect the intrinsic imprecision of the potential responses, combined with a free response format which allows us to reflect the inherent subjectivity of these responses. In this way, the variability and diversity are exploited more accurately.

In this respect, an example of such a survey has been carried out during the II Summer School of the European Centre for Soft Computing (Mieres, Spain) in July 2008. For each course, students attending it (who are familiar with fuzzy numbers because of the courses belonging to a specialized teaching program) have been inquired to represent their opinion/valuation about 5 different aspects of each course. Since opinion/valuation assessments are intrinsically imprecise, students have been requested to reply by using fuzzy

Q1. Motivation of the course

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Q2. Intellectual challenge provided by the course

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Q3. Lecturer performance

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Q4. Quality of the course material

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Q5. Overall rating

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Fig. 1. Free fuzzy numbered response format questionnaire for each student and course

numbers on $[0, 100]$ (as the set of values which are assumed to be potentially compatible with any possible response), with 0 and 100 representing the lowest and greatest valuation, respectively. Figure 1 shows a template of the form that the students have filled out.

To ease the drawing of the fuzzy numbers the use of trapezoidal numbers $\text{Tra}(i_0, i_1, s_1, s_0)$ has been suggested. Figure 2 shows the form as filled out by one of the students attending one of the courses in the above mentioned Summer School.

Guidelines on the fuzzy assessments and interpretations have been indicated in González-Rodríguez *et al.* [16]. As it has already been commented, one of the obvious advantages of these fuzzy numbered response format questionnaires is that they allow full freedom in describing valuations and judgements, high expressiveness, flexibility and accuracy to state them, and they capture high variability and subjectivity (for instance only three coincidences have been detected in the responses to the ‘overall rating’ of the 27 students attending one of the courses, for which data have been collected in Table 2).

The fuzzy median associated with the data in Table 2 is depicted in Figure 3 in which, as indicated in Remark 2, the median (which will be uniquely defined without need to apply any convention) does not correspond to any of the data.

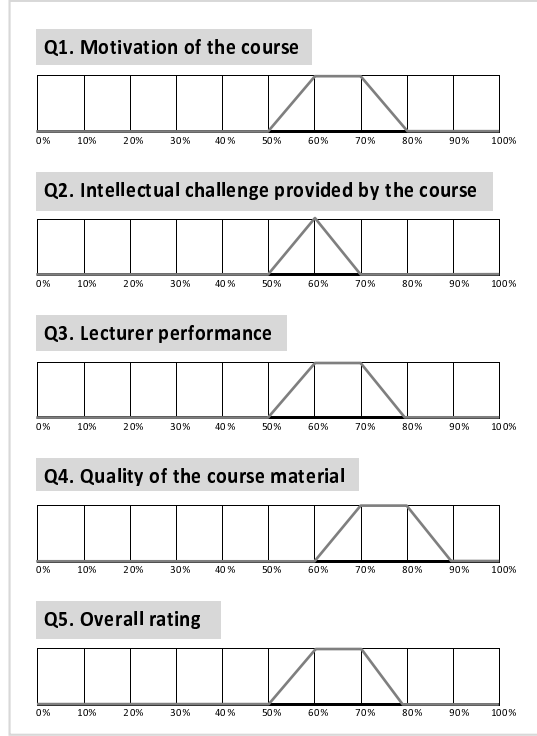


Fig. 2. Example of the responses supplied by a concrete student for a given course

St	i_0	i_1	s_1	s_0	St	i_0	i_1	s_1	s_0	St	i_0	i_1	s_1	s_0
1	50	60	70	78	10	80	85	85	90	19	57	60	64	67
2	33	36	50	57	11	60	70	70	80	20	40	50	50	60
3	44	50	70	77	12	30	50	50	80	21	30	40	40	50
4	84	88	94	97	13	50	60	70	80	22	65	70	80	85
5	50	60	70	80	14	50	60	60	70	23	80	86	94	100
6	50	60	70	80	15	57	66	74	100	24	80	90	90	100
7	35	45	55	66	16	4	6	12	20	25	58	65	74	78
8	67	73	77	80	17	60	70	90	100	26	60	70	80	90
9	60	65	65	70	18	65	70	75	80	27	60	89	89	99

Table 2. Trapezoidal responses to the ‘overall rating’ of 27 students of a course

On the basis of this example one can get a first illustration of the idea that the mean is less robust than the median (this assertion being formally and empirically supported later). Thus, if the answer of the 16th student (which clearly represents an outlier in the sample) is removed from the dataset, the median scarcely varies, whereas the mean ‘increases’ around over 2 units (more precisely, the mean answer for the 27 students is $Tra(54.037, 62.740, 69.185, 78.296)$, whereas once the 16th answer is removed the mean equals $Tra(55.962, 64.923, 71.385, 80.539)$).

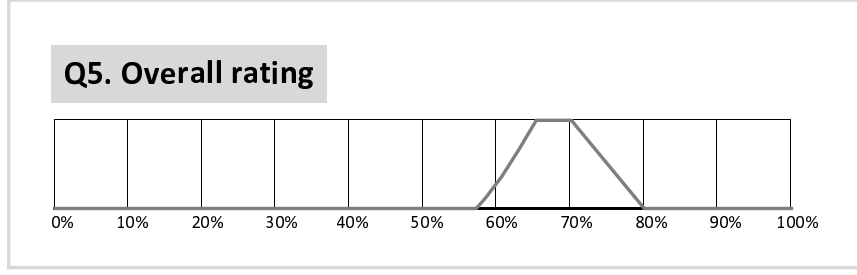


Fig. 3. Median of the RFN ‘overall rating’ of the course in Example 4.1

Generally speaking, questionnaires like the one in this example provide richer and more variable and diverse information than traditional ones, and statistics based on this information make more sense and will be more informative.

5 Basic properties of the median of a random fuzzy number

The median of an RFN preserves most of the basic properties of the median of a random variable. Thus, it can be verified that

Proposition 5.1 *$\widetilde{\text{Me}}$ is equivariant by ‘linear’ transformations, that is, if $\gamma \in \mathbb{R}$, $\tilde{U} \in \mathcal{F}_c^*(\mathbb{R})$ and \mathcal{X} is an RFN, then*

$$\widetilde{\text{Me}}(\gamma \cdot \mathcal{X} + \tilde{U}) = \gamma \cdot \widetilde{\text{Me}}(\mathcal{X}) + \tilde{U}.$$

Consequently, if \mathcal{X} is an RFN associated with the probability space (Ω, \mathcal{A}, P) and the distribution of \mathcal{X} is degenerate at a fuzzy number $\tilde{U} \in \mathcal{F}_c^(\mathbb{R})$ (i.e., $\mathcal{X} = \tilde{U}$ a.s. $[P]$), then $\widetilde{\text{Me}}(\mathcal{X}) = \tilde{U}$.*

In Section 4 we have outlined that the median has been extended by considering it as a value minimizing the mean distance to the distribution of the random element through an L^1 -type metric, since fuzzy numbers cannot be ranked through a universally acceptable total ordering. However, it should be emphasized that the median of an RFN as introduced in this paper can be formalized as a ‘middle position’ value with respect to the *fuzzy max partial order*, whenever this order applies. The fuzzy max order on $\mathcal{F}_c^*(\mathbb{R})$ was introduced by Dubois and Prade [10], and equivalent definitions were stated by Ramík and Římánek [32] and more recently by Valvis [36]. It is the natural levelwise extension through the support function (i.e., through the inf / sup characterization) of the product order on \mathbb{R}^2 , so that $\tilde{U} \lesssim \tilde{V}$ if and only if for all $\alpha, \lambda \in [0, 1]$ one has that $\lambda \sup \tilde{U}_\alpha + (1 - \lambda) \inf \tilde{U}_\alpha \leq \lambda \sup \tilde{V}_\alpha + (1 - \lambda) \inf \tilde{V}_\alpha$. The main drawback for this ranking is the fact that it only leads to a partial ordering and many fuzzy numbers cannot be compared with it. However, it is often viewed as a quite acceptable ranking criterion which is considered as a

pattern which should be preserved for any more widely applicable or complete suggested ranking.

Proposition 5.2 *For any sample or finite population $(\omega_1, \dots, \omega_n)$ for which the values of an RFN \mathcal{X} satisfy that*

$$\mathcal{X}(\omega_1) \lesssim \dots \lesssim \mathcal{X}(\omega_n)$$

we have that

- *if n is odd, then*

$$\widetilde{\text{Me}}(\mathcal{X}) = \mathcal{X}(\omega_{(n+1)/2}),$$

- *if n is even, then*

$$\widetilde{\text{Me}}(\mathcal{X}) = \frac{1}{2} \cdot (\mathcal{X}(\omega_{n/2}) + \mathcal{X}(\omega_{(n/2)+1})).$$

6 Consistency and robustness of the sample median and comparisons with the sample mean

The inferential behavior of the median of an RFN will now be analyzed. In this respect, the ρ_1 -strong consistency and the finite sample breakdown point to examine its robustness will be now discussed. Similarly as for the real-valued case, under mild conditions the sample median is a *strongly consistent estimator* of the population median, that is,

Proposition 6.1 *Let \mathcal{X} be an RFN associated with a probability space (Ω, \mathcal{A}, P) satisfying that $\text{Me}(\inf \mathcal{X}_\alpha)$ and $\text{Me}(\sup \mathcal{X}_\alpha)$ are actually unique (i.e., they are unique without applying the convention in Theorem 4.1) for each α .*

If $\widehat{\text{Me}}(\mathcal{X})_n$ denotes the sample median corresponding to a simple random sample $(\mathcal{X}_1, \dots, \mathcal{X}_n)$ from \mathcal{X} (i.e., $\mathcal{X}_1, \dots, \mathcal{X}_n$ are independent and identically distributed as \mathcal{X}), then we have that

$$\lim_{n \rightarrow \infty} \rho_1(\widehat{\text{Me}}(\mathcal{X})_n, \widetilde{\text{Me}}(\mathcal{X})) = 0 \quad \text{a.s. } [P].$$

Proof. We have that,

$$\begin{aligned} & P \left(\lim_{n \rightarrow \infty} \rho_1(\widehat{\text{Me}}(\mathcal{X})_n, \widetilde{\text{Me}}(\mathcal{X})) = 0 \right) \\ &= P \left(\lim_{n \rightarrow \infty} \left(\frac{1}{2} \int_{(0,1]} |\widehat{\text{Me}}(\inf \mathcal{X}_\alpha)_n - \widetilde{\text{Me}}(\inf \mathcal{X}_\alpha)| d\alpha \right. \right. \\ & \quad \left. \left. + \frac{1}{2} \int_{(0,1]} |\widehat{\text{Me}}(\sup \mathcal{X}_\alpha)_n - \widetilde{\text{Me}}(\sup \mathcal{X}_\alpha)| d\alpha \right) = 0 \right) \\ &\geq P \left(\left(\lim_{n \rightarrow \infty} \sup_{\alpha \in (0,1]} |\widehat{\text{Me}}(\inf \mathcal{X}_\alpha)_n - \widetilde{\text{Me}}(\inf \mathcal{X}_\alpha)| = 0 \right) \right) \end{aligned}$$

$$\cap \left(\lim_{n \rightarrow \infty} \sup_{\alpha \in (0,1]} |\widehat{\text{Me}(\sup \mathcal{X}_\alpha)}_n - \widehat{\text{Me}(\sup \mathcal{X}_\alpha)}| = 0 \right)$$

For the considered sample of fuzzy data, whatever $n \in \mathbb{N}$ may be there exists an $\alpha_0 \in (0,1]$ such that $0 < \sup_{\alpha \in (0,1]} |\widehat{\text{Me}(\inf \mathcal{X}_\alpha)}_n - \widehat{\text{Me}(\inf \mathcal{X}_\alpha)}| < |\widehat{\text{Me}(\inf \mathcal{X}_{\alpha_0})}_n - \widehat{\text{Me}(\inf \mathcal{X}_{\alpha_0})}| + 1/n$, whence by taking limits and applying the continuity of the absolute value function and the probability

$$\begin{aligned} & P \left(\lim_{n \rightarrow \infty} \sup_{\alpha \in (0,1]} |\widehat{\text{Me}(\inf \mathcal{X}_\alpha)}_n - \widehat{\text{Me}(\inf \mathcal{X}_\alpha)}| = 0 \right) \\ & \geq P \left(\lim_{n \rightarrow \infty} |\widehat{\text{Me}(\inf \mathcal{X}_{\alpha_0})}_n - \widehat{\text{Me}(\inf \mathcal{X}_{\alpha_0})}| = 0 \right) \\ & = P \left(\lim_{n \rightarrow \infty} (\widehat{\text{Me}(\inf \mathcal{X}_{\alpha_0})}_n - \widehat{\text{Me}(\inf \mathcal{X}_{\alpha_0})}) = 0 \right). \end{aligned}$$

Under the assumption of uniqueness for the median of $\inf \mathcal{X}_{\alpha_0}$, the sample median is a strongly consistent estimator of the population median, and hence

$$P \left(\lim_{n \rightarrow \infty} (\widehat{\text{Me}(\inf \mathcal{X}_{\alpha_0})}_n - \widehat{\text{Me}(\inf \mathcal{X}_{\alpha_0})}) = 0 \right) = 1.$$

By following similar arguments, one can prove that whatever $n \in \mathbb{N}$ may be there exists $\alpha'_0 \in (0,1]$ such that

$$P \left(\lim_{n \rightarrow \infty} (\widehat{\text{Me}(\sup \mathcal{X}_{\alpha'_0})}_n - \widehat{\text{Me}(\sup \mathcal{X}_{\alpha'_0})}) = 0 \right) = 1.$$

Consequently,

$$P \left(\lim_{n \rightarrow \infty} \rho_1(\widehat{\text{Me}(\mathcal{X})}_n, \widetilde{\text{Me}(\mathcal{X})}) = 0 \right) = 1. \quad \square$$

Let us now discuss the robustness of the sample median of an RFN as an estimator of the population median in contrast to that of the sample mean of an RFN as an estimator of the population mean.

Before presenting a formal discussion and comparison, we analyze the simulations in Section 3 when the mean is replaced by the median. As for Table 1, to determine the effect of the contamination on the median of the RFN \mathcal{X} , the expected distance between the noncontaminated ‘distribution’ and the Monte Carlo approximated median is collected in Table 3 for the different values of c_p and C_D and the Cases 1 and 2 in Table 1. Contrary to the results in Table 1, the results in Table 3 show that the expected distance between the noncontaminated distribution and the sample median only slightly changes when the amount of contamination is increased, even when the contamination lies far from the noncontaminated distribution.

The analysis of the robustness of the median in comparison to the mean is now made through the so-called *finite sample breakdown point* (fsbp for short),

c_P	c_D	Case 1 (ρ_1)	Case 1 (ρ_2)	Case 2 (ρ_1)	Case 2 (ρ_2)
0.0	0	1.387025	1.553613	1.395168	1.602187
0.0	1	1.381609	1.548223	1.395386	1.602493
0.0	5	1.387965	1.554998	1.396987	1.606656
0.0	10	1.381480	1.547408	1.396204	1.604439
0.0	100	1.387858	1.554551	1.390518	1.597395
0.1	0	1.390256	1.563623	1.394283	1.602143
0.1	1	1.394385	1.569681	1.387921	1.595771
0.1	5	1.400976	1.575503	1.400777	1.611329
0.1	10	1.400868	1.575568	1.398967	1.609700
0.1	100	1.400570	1.577209	1.399620	1.612492
0.2	0	1.414179	1.595659	1.396819	1.605635
0.2	1	1.431032	1.617461	1.403823	1.613799
0.2	5	1.450967	1.639115	1.429443	1.644502
0.2	10	1.438225	1.625088	1.442132	1.658978
0.2	100	1.451449	1.639004	1.453777	1.671996
0.4	0	1.587835	1.795133	1.379690	1.586880
0.4	1	1.731092	1.950794	1.444821	1.663506
0.4	5	1.947556	2.176999	1.774923	2.038919
0.4	10	2.022447	2.250256	1.886229	2.147031
0.4	100	2.072649	2.288222	2.067098	2.291444

Table 3. Mean distances of the mixed (partially contaminated and noncontaminated) sample median to the noncontaminated distribution of an RFN

quantifying the minimum proportion of sample data which should be perturbed to get an arbitrarily large or small estimator value. Following Donoho and Huber the fsbp of the sample median in a sample of size n from an RFN \mathcal{X} is given by

$$\begin{aligned} & \text{fsbp}(\widehat{\text{Me}(\mathcal{X})}_n, \tilde{\mathbf{x}}_n, \rho_1) \\ &= \frac{1}{n} \min \left\{ k \in \{1, \dots, n\} : \sup_{Q_{n,k}} \rho_1(\widehat{\text{Me}(P_n)}, \widehat{\text{Me}(Q_{n,k})}) = \infty \right\}, \end{aligned}$$

where $\tilde{\mathbf{x}}_n$ denotes the considered sample of n data from the metric space $(\mathcal{F}_c^*(\mathbb{R}), \rho_1)$ in which $\sup_{\tilde{U}, \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})} \rho_1(\tilde{U}, \tilde{V}) = \infty$, P_n is the empirical distribution of $\tilde{\mathbf{x}}_n$ and $Q_{n,k}$ is the empirical distribution of sample $\tilde{\mathbf{y}}_{n,k}$ obtained from the original one $\tilde{\mathbf{x}}_n$ by perturbing at most k components. Then, we have that

Proposition 6.2 *The finite sample breakdown point of the sample median from an RFN \mathcal{X} , $\text{fsbp}(\widehat{\text{Me}(\mathcal{X})}_n)$, equals*

$$\text{fsbp}(\widehat{\text{Me}(\mathcal{X})}_n) = \frac{1}{n} \cdot \lfloor \frac{n+1}{2} \rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the floor function.

Proof. First note that the condition $\sup_{\tilde{U}, \tilde{V} \in \mathcal{F}_c^*(\mathbb{R})} \rho_1(\tilde{U}, \tilde{V}) = \infty$ is satisfied in this case, since $\rho_1(\mathbf{1}_{[n-1, n+1]}, \mathbf{1}_{[-n-1, -n+1]}) = 2n$.

Furthermore,

$$\begin{aligned} \rho_1(\widehat{\text{Me}}(P_n), \widehat{\text{Me}}(Q_{n,k})) &\geq \int_{(0,1]} \frac{1}{2} \cdot |\inf(\widehat{\text{Me}}(P_n))_\alpha - \inf(\widehat{\text{Me}}(Q_{n,k}))_\alpha| d\alpha \\ &= \int_{(0,1]} \frac{1}{2} \cdot |\text{Me}(\inf(P_n)_\alpha) - \text{Me}(\inf(Q_{n,k})_\alpha)| d\alpha. \end{aligned}$$

Therefore, by recalling the fsbp for the sample median of a real-valued random variable, one can conclude that whenever at least $\lfloor \frac{n+1}{2} \rfloor$ elements $\tilde{x}_i \in \mathcal{F}_c^*(\mathbb{R})$ of $\tilde{\mathbf{x}}_n$ are replaced by other arbitrarily ‘large’ elements in $\mathcal{F}_c^*(\mathbb{R})$ so that

$$\sup_{Q_{n,k}} \int_{(0,1]} \frac{1}{2} \cdot |\text{Me}(\inf(P_n)_\alpha) - \text{Me}(\inf(Q_{n,k})_\alpha)| d\alpha = \infty,$$

we have that

$$\begin{aligned} &\sup_{Q_{n,k}} \rho_1(\widehat{\text{Me}}(P_n), \widehat{\text{Me}}(Q_{n,k})) \\ &\geq \frac{1}{2} \cdot \sup_{Q_{n,k}} \int_{(0,1]} \frac{1}{2} \cdot |\text{Me}(\inf(P_n)_\alpha) - \text{Me}(\inf(Q_{n,k})_\alpha)| d\alpha = \infty, \end{aligned}$$

whence

$$\text{fsbp}(\widehat{\text{Me}}(\mathcal{X})_n, \tilde{\mathbf{x}}_n, \rho_1) \leq \frac{1}{n} \cdot \lfloor \frac{n+1}{2} \rfloor.$$

On the other hand, by using the fsbp for the sample median of a real-valued random variable, we have that for all α

$$\begin{aligned} \min \left\{ k \in \{1, \dots, n\} : \sup_{Q_{n,k}} |\text{Me}(\inf(P_n)_\alpha) - \text{Me}(\inf(Q_{n,k})_\alpha)| = \infty \right\} &= \lfloor \frac{n+1}{2} \rfloor, \\ \min \left\{ k \in \{1, \dots, n\} : \sup_{Q_{n,k}} |\text{Me}(\sup(P_n)_\alpha) - \text{Me}(\sup(Q_{n,k})_\alpha)| = \infty \right\} &= \lfloor \frac{n+1}{2} \rfloor, \end{aligned}$$

whence

$$\begin{aligned} \sup_{Q_{n, \lfloor \frac{n+1}{2} \rfloor - 1}} |\text{Me}(\inf(P_n)_\alpha) - \text{Me}(\inf(Q_{n, \lfloor \frac{n+1}{2} \rfloor - 1})_\alpha)| &= M_1 < \infty, \\ \sup_{Q_{n, \lfloor \frac{n+1}{2} \rfloor - 1}} |\text{Me}(\sup(P_n)_\alpha) - \text{Me}(\sup(Q_{n, \lfloor \frac{n+1}{2} \rfloor - 1})_\alpha)| &= M_2 < \infty, \end{aligned}$$

and therefore

$$\begin{aligned} &\sup_{Q_{n, \lfloor \frac{n+1}{2} \rfloor - 1}} \rho_1(\widehat{\text{Me}}(P_n), \widehat{\text{Me}}(Q_{n, \lfloor \frac{n+1}{2} \rfloor - 1})) \\ &= \sup_{Q_{n, \lfloor \frac{n+1}{2} \rfloor - 1}} \left[\frac{1}{2} \int_{(0,1]} |\text{Me}(\inf(P_n)_\alpha) - \text{Me}(\inf(Q_{n, \lfloor \frac{n+1}{2} \rfloor - 1})_\alpha)| d\alpha \right] \end{aligned}$$

$$\left. + \frac{1}{2} \int_{(0,1]} |\text{Me}(\widehat{\inf(P_n)_\alpha}) - \text{Me}(\widehat{\inf(Q_{n, \lfloor \frac{n+1}{2} \rfloor - 1})_\alpha})| d\alpha \right] \leq \frac{M_1 + M_2}{2} < \infty.$$

Consequently,

$$\min \left\{ k \in \{1, \dots, n\} : \sup_{Q_{n,k}} \rho_1(\widehat{\text{Me}(P_n)}, \widehat{\text{Me}(Q_{n,k})}) = \infty \right\} > \lfloor \frac{n+1}{2} \rfloor - 1,$$

whence

$$\text{fsbp}(\widehat{\text{Me}(\mathcal{X})}_n, \tilde{\mathbf{x}}_n, \rho_1) \geq \frac{1}{n} \cdot \lfloor \frac{n+1}{2} \rfloor. \quad \square$$

The fsbp can be also computed for the sample mean of an RFN and, by comparing it with that for the median. Thus,

Theorem 6.3 *The finite sample breakdown point of the sample mean from an RFN \mathcal{X} , $\text{fsbp}(\overline{\mathcal{X}}_n)$, is lower than that for the sample median for sample sizes $n > 2$.*

Proof. Indeed, by arguing like for the preceding proposition we have that

$$\text{fsbp}(\overline{\mathcal{X}}_n, \tilde{\mathbf{x}}_n, \rho_1) = \frac{1}{n},$$

and, consequently,

$$\text{fsbp}(\widehat{\text{Me}(\mathcal{X})}_n, \tilde{\mathbf{x}}_n, \rho_1) \geq \frac{n/2}{n} = \frac{1}{2} > \frac{1}{n} = \text{fsbp}(\overline{\mathcal{X}}_n, \tilde{\mathbf{x}}_n, \rho_1). \quad \square$$

The sample mean has the lowest possible breakdown point while the sample median can withstand up to 50% of contamination. This huge difference can be also stressed in the fuzzy case. It means that the definition of fuzzy median in this paper succeeds in inheriting the robustness properties of the real valued sample median.

The theoretical conclusion in Theorem 6.3 can be corroborated empirically by analyzing the simulations in Section 3 and those at the beginning of this section. Moreover, and on the basis of these simulations an additional table has been constructed. Table 4 gathers empirical results for the influence of contamination on both the sample mean and median, by computing the distances between the mean/median of the noncontaminated sample and the mean/median of the contaminated sample, respectively, for the different values of c_p and C_D and the Cases 1 (C1) and 2 (C2) in Tables 1 and 3.

On the basis of these simulations and by comparing Tables 1 and 3, and the results in Table 4, one can empirically conclude that

c_P	c_D	Case 1				Case 2			
		means		medians		means		medians	
		ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2	ρ_1	ρ_2
0.0	0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.0	1	0.005620	0.006591	0.009937	0.010690	0.006110	0.008218	0.004619	0.005502
0.0	5	0.006565	0.006644	0.007940	0.008527	0.010661	0.012205	0.004700	0.005241
0.0	10	0.005362	0.005994	0.006942	0.008223	0.007431	0.008902	0.005293	0.005719
0.0	100	0.004541	0.004918	0.004439	0.004863	0.005423	0.006698	0.003916	0.004392
0.1	0	0.451078	0.459370	0.158807	0.162013	0.054203	0.055236	0.005964	0.006834
0.1	1	0.601447	0.620249	0.192335	0.196861	0.106741	0.137490	0.046193	0.059756
0.1	5	1.200294	1.321521	0.212689	0.216379	0.677751	0.847402	0.123114	0.138903
0.1	10	1.959302	2.227514	0.219611	0.222192	1.416000	1.750975	0.154771	0.167474
0.1	100	15.453551	18.647596	0.225127	0.226117	14.911972	18.185063	0.173094	0.177625
0.2	0	0.897153	0.913668	0.352701	0.359852	0.097041	0.098726	0.008799	0.011117
0.2	1	1.214963	1.252698	0.437944	0.447316	0.218693	0.286074	0.123952	0.155375
0.2	5	2.407951	2.647567	0.486023	0.494771	1.365394	1.708745	0.299863	0.344031
0.2	10	3.896666	4.443769	0.489210	0.495759	2.853653	3.521330	0.380098	0.411177
0.2	100	30.891914	37.284770	0.515718	0.519286	29.845030	36.393312	0.433573	0.443462
0.4	0	1.792771	1.826105	0.936099	0.961331	0.193307	0.196715	0.016402	0.020866
0.4	1	2.398392	2.475396	1.218930	1.261627	0.439475	0.576132	0.333567	0.431527
0.4	5	4.798148	5.278300	1.599580	1.650905	2.741171	3.426632	1.089851	1.300477
0.4	10	7.811155	8.902006	1.732302	1.776734	5.698655	7.040080	1.345853	1.483912
0.4	100	61.796235	74.580521	1.813070	1.830480	59.694633	72.784278	1.696111	1.731362

Table 4. Distances between the sample mixed (partially contaminated and noncontaminated) mean/median to the noncontaminated one for an RFN

- Obviously, in case $c_P = 0$ we have that the mean ρ_1 -distance w.r.t. sample median is lower than w.r.t. the sample mean, whereas the mean ρ_2 -distance w.r.t. sample median is greater than w.r.t. the sample mean.
- For a fixed level of contamination c_P , the farther the contaminated distribution from the noncontaminated one the substantially greater mean (both, ρ_1 and ρ_2) distance between the approximated mean and the noncontaminated distribution, whereas for the approximated median the increase is modest; actually this mean distance asymptotically would only depend on a certain fractile of the noncontaminated distribution.
- For a fixed level of contamination c_P , the farther the contaminated distribution from the noncontaminated one, the substantially greater distance between the contaminated and the noncontaminated means, whereas for the medians the increase is not really substantial.

7 Concluding remarks

Random fuzzy numbers are a well-stated and supported tool to model and handle random elements taking on fuzzy numbered-valued data. They fit and apply to many fields like Social and Behavioral Sciences (see, for instance, Smithson [33]), Medicine (see, for instance, Hu *et al.* [18]) or Fuzzy Control (see, for instance, Faraz and Shapiro [11]). Actually, the last mentioned paper motivates also the interest of summarizing the centrality of a sample of fuzzy data in control charts (especially in cases the sample is asymmetric) in a way such that data are monitored as fuzzy numbers instead of monitored as real-valued ones after a defuzzification process.

This paper has explored the notion of median of an RFN on the basis of an L^1 -type metric between fuzzy numbers. Since we present an introductory

work on the topic, there are many open problems of immediate interest to be examined, namely,

- To formalize and develop comparative studies with other approaches for the median of an RFN, as the one based on the functional identification of fuzzy numbers and the corresponding induced functional median (cf. Fraiman and Muñiz [12], Cuevas *et al.* [7], Gervini [13], etc.).
- To consider other L^1 metrics (like those based on the mid/spread representation of fuzzy numbers (see Trutschnig *et al.* [34]) and formalize the median in a way similar as in the paper; the main problems which can arise are those associated with either the difficulties to guarantee the existence of a fuzzy number minimizing the mean distance or to find appropriate conventions to get them. Alternatively, in case these difficulties cannot be easily overcome, it could be convenient to follow approximation ideas by Luenberger [24] through the support function.
- To formalize and examine properties of the mean ρ_1 -distance of the distribution of \mathcal{X} to $\widetilde{\text{Me}}(\mathcal{X})$, as a measure of the average dispersion w.r.t. the median.

To conclude the paper we wish to point out that the notion of median in an imprecise setting by Couso and Sánchez [6] refer to a completely different setting and approach which is based on imprecise probabilities instead of imprecise data. In general, for most of the situations and developments, the concepts, tools and methods for this approach/setting don't make sense or cannot be applied under the available information and assumptions for the one in this paper, and conversely. The median in [6] has been introduced as a definition based on some rankings or differences, which are well-defined in the approach involving imprecise probabilities; it is in fact defined as a middle/intermediate position measure, which makes sense in this setting. In contrast to this notion, the median in this paper has been introduced as a measure minimizing a mean L^1 distance (i.e., as a central tendency measure instead of a middle position one). In the setting of random fuzzy sets, as outlined along the paper, neither the ranking nor the differences are well-defined in general. Actually, the particular and easy-to-handle choice suggested in Theorem 4.1 is not presented as a definition, but it has been obtained as an easy-to-use solution minimizing the mean distance. To guarantee this solution is in fact a fuzzy number both the considered convention and metric have been crucial, because of the lack of linearity of the median operator for real-valued random variables.

References

- [1] Bouchon-Meunier, B., Kosheleva, O., Kreinovich, V., Nguyen, H.T., 1997. Fuzzy numbers are the only fuzzy sets that keep invertible operations invertible. *Fuzzy Sets and Systems* 91, 155-163.

- [2] Castaing, C., Valadier, M., 1977. Convex Analysis and Measurable Multifunctions. Lec. Notes in Math. 580. Springer-Verlag, Berlin.
- [3] Colubi, A., Domínguez-Menchero, J.S., López-Díaz, M., Ralescu, D.A., 2001. On the formalization of fuzzy random variables. Inform. Sci. 133, 3–6.
- [4] Colubi, A., Domínguez-Menchero, J. S., López-Díaz, M., Ralescu, D. A., 2002. A $D_E[0, 1]$ -representation of random upper semicontinuous functions. Proc. Am. Math. Soc. 130, 3237–3242.
- [5] Colubi, A., López-Díaz, M., Domínguez-Menchero, J.S., Gil, M.A., 1999. A generalized strong law of large numbers. Prob. Theor. Rel. Fields 114, 401–417.
- [6] Couso, I., Sánchez, L., 2010. The behavioral meaning of the median. In: Combining Soft Computing and Statistical Methods in Data Analysis (Borgelt, C., González-Rodríguez, G., Trutschnig, W., Gil, M.A., Grzegorzewski, P., Hryniewicz, O., eds.). Springer, Heidelberg, 115–123.
- [7] Cuevas, A., Febrero, M., Fraiman, R., 2006. On the use of the bootstrap for estimating functions with functional data. Comp. Stat. Data Anal. 51, 1063–1074.
- [8] Diamond, P., Kloeden, P., 1999. Metric spaces of fuzzy sets. Fuzzy Sets and Systems 100, 63–71.
- [9] Donoho, D.L., Huber, P.J., 1983. The notion of breakdown point. In: A Festschrift for Erich L. Lehmann (Bickel, P.J., Doksum, K., Hodges, J.L. Jr. eds.). Wadsworth, Belmont, 157–184.
- [10] Dubois, D., Prade, H., 1980. Systems of linear fuzzy constraints. Fuzzy Sets and Systems 3, 37–48.
- [11] Faraz, A., Shapiro, A.F., 2010. An application of fuzzy random variables to control charts. Fuzzy Sets and Systems 161, 2684–2694.
- [12] Fraimann, R., Muñiz, G., 2001. Trimmed means for functional data. Test 10, 419–440.
- [13] Gervini, D., 2008. Robust functional estimation using the spatial median and spherical principal components. Biometrika 95, 587–600.
- [14] Gil, M.A., Montenegro, M., González-Rodríguez, G., Colubi, A., Casals, M.R., 2006. Bootstrap approach to the multi-sample test of means with imprecise data. Comp. Stat. Data Anal. 51, 148–162.
- [15] González-Rodríguez, G., Colubi, A., Gil, M.A., 2006a. A fuzzy representation of random variables: an operational tool in exploratory analysis and hypothesis testing. Comp. Stat. Data Anal. 51, 163–176.
- [16] González-Rodríguez, G., Colubi, A., Gil, M.A., 2010. Fuzzy data treated as functional data. A one-way ANOVA test approach. Comp. Stat. Data Anal. In press (doi:10.1016/j.csda.2010.06.013).

- [17] González-Rodríguez, G., Montenegro, M., Colubi, A., Gil, M.A., 2006b. Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data. *Fuzzy Sets and Systems* 157, 2608–2613.
- [18] Hu, H-Y., Lee, Y-C., Yen, T-M., 2010. Service quality gaps analysis based on Fuzzy linguistic SERVQUAL with a case study in hospital out-patient services. *The TQM Journal* 22, 499–515.
- [19] Hukuhara, M., 1967. Intégration des applications mesurables dont la valeur est un compact convexe. *Funkcial. Ekvac.* 10, 205–223.
- [20] Klement, E.P., Puri, M.L., Ralescu, D.A., 1986. Limit theorems for fuzzy random variables. *Proc. R. Soc. Lond. A* 407, 171–182.
- [21] Körner, R., 2000. An asymptotic α -test for the expectation of random fuzzy variables. *J. Stat. Plann. Inference* 83, 331–346.
- [22] Körner, R., Näther, W., 2002. On the variance of random fuzzy variables. In: *Statistical Modeling, Analysis and Management of Fuzzy Data* (Bertoluzza, C., Gil, M.A., Ralescu, D.A. eds.). Physica-Verlag, Heidelberg, 22–39.
- [23] Li, S., Ogura, Y., 2006. Strong laws of large numbers for independent fuzzy set-valued random variables. *Fuzzy Sets and Systems* 157, 2569–2578.
- [24] Luenberger, K., 1968. *Optimization by Vector Space Methods*. Wiley, New York.
- [25] Molchanov, I., 1999. On strong laws of large numbers for random upper semicontinuous functions. *J. Math. Anal. Appl.* 235, 349–355.
- [26] Montenegro, M., Casals, M. R., Lubiano, M. A., Gil, M. A., 2001. Two-sample hypothesis tests of means of a fuzzy random variable. *Inform. Sci* 133, 89–100.
- [27] Montenegro, M., Colubi, A., Casals, M. R., Gil, M. A., 2004. Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika* 59, 31–49.
- [28] Phillis, Y.A., Kouikoglou, V.S., 2009. *Fuzzy Measurement of Sustainability*. Nova Sci. Pub., New York.
- [29] Proske, F.N., Puri, M.L., 2003. A strong law of large numbers for generalized random sets from the viewpoint of empirical processes. *Proc. Am. Math. Soc.* 131, 2937–2944.
- [30] Puri, M.L., Ralescu, D.A., 1985. The concept of normality for fuzzy random variables. *Ann. Probab.* 11, 1373–1379.
- [31] Puri, M.L., Ralescu, D.A. 1986. Fuzzy random variables. *J. Math. Anal. Appl.* 114, 409–422.
- [32] Ramík, J., Římánek, J., 1985. Inequality relation between fuzzy numbers and its use in fuzzy optimization. *Fuzzy Sets and Systems* 16, 123–138.
- [33] Smithson, M., 1982. Applications of Fuzzy Set concepts to Behavioral Sciences. *Math. Soc. Sci.* 2, 257–274. *Inform. Sci.* 179, 3964–3972.

- [34] Trutschnig, W., González-Rodríguez, G., Colubi, A., Gil, M.A., 2009. A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. Inform. Sci. 179, 3964–3972.
- [35] Trutschnig, W., Lubiano, M.A., 2010. SAFD: Statistical Analysis of Fuzzy Data (R package) (<http://bellman.ciencias.uniovi.es/SMIRE/SAFDpackage.html>).
- [36] Valvis, E., 2009. A new linear ordering of fuzzy numbers on subsets of $\mathcal{F}(\mathbb{R})$. Fuzzy Optim. Decis. Making 8, 141–163.
- [37] Yager, R.R., 1981. A Procedure for Ordering Fuzzy Subsets of the Unit Interval. Inform. Sci. 24, 143–161.
- [38] Zadeh, L.A., 1975. The concept of a linguistic variable and its application to approximate reasoning, Part 1. Inform. Sci. 8, 199–249; Part 2. Inform. Sci. 8, 301–353; Part 3. Inform. Sci. 9, 43–80.